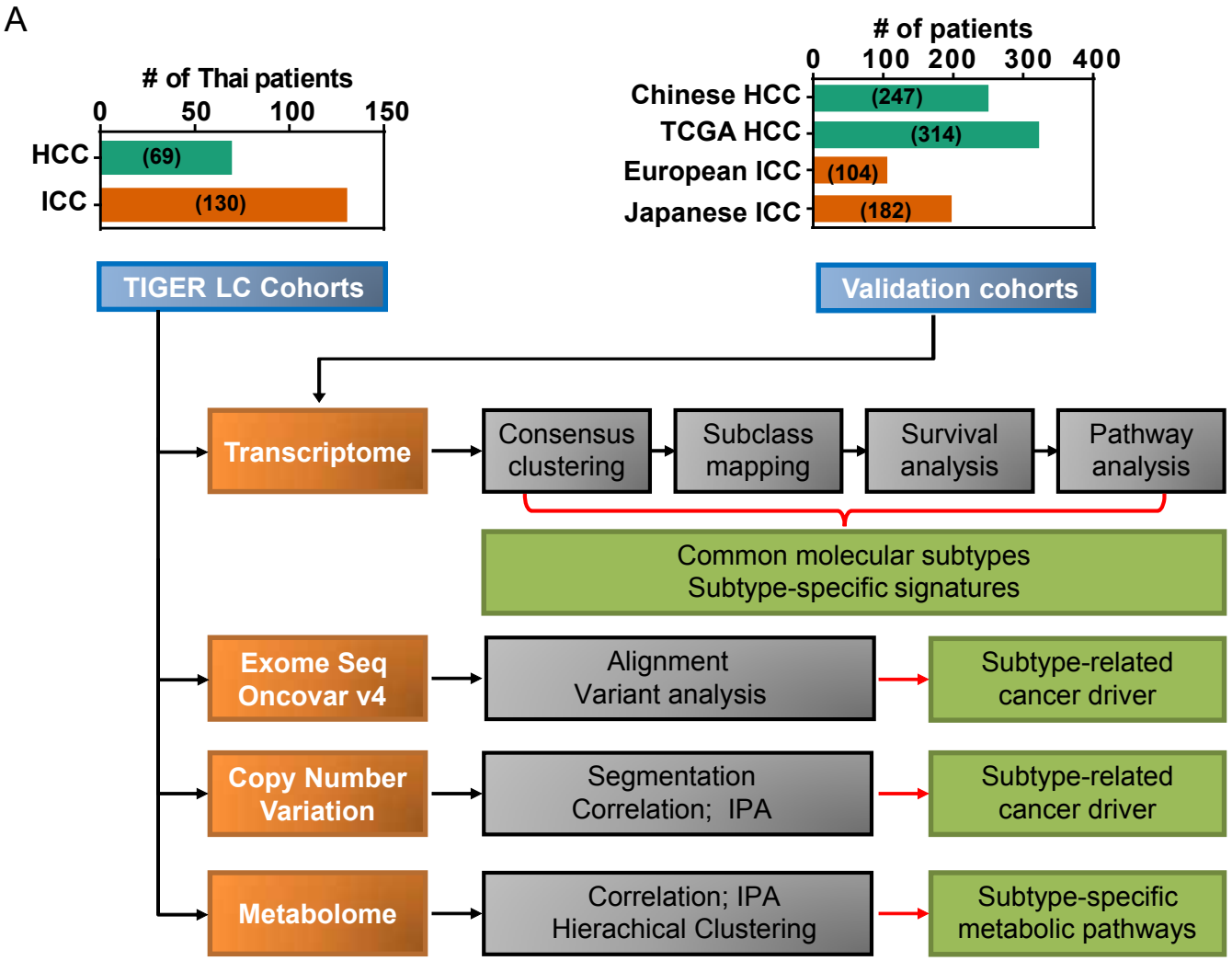
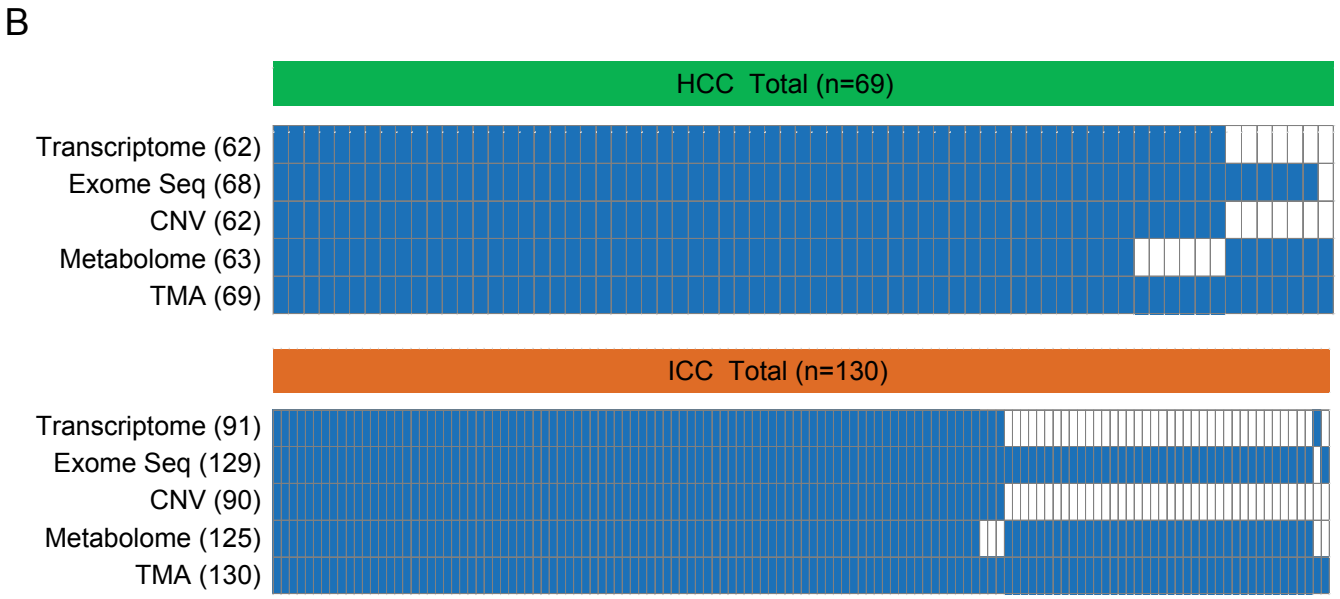
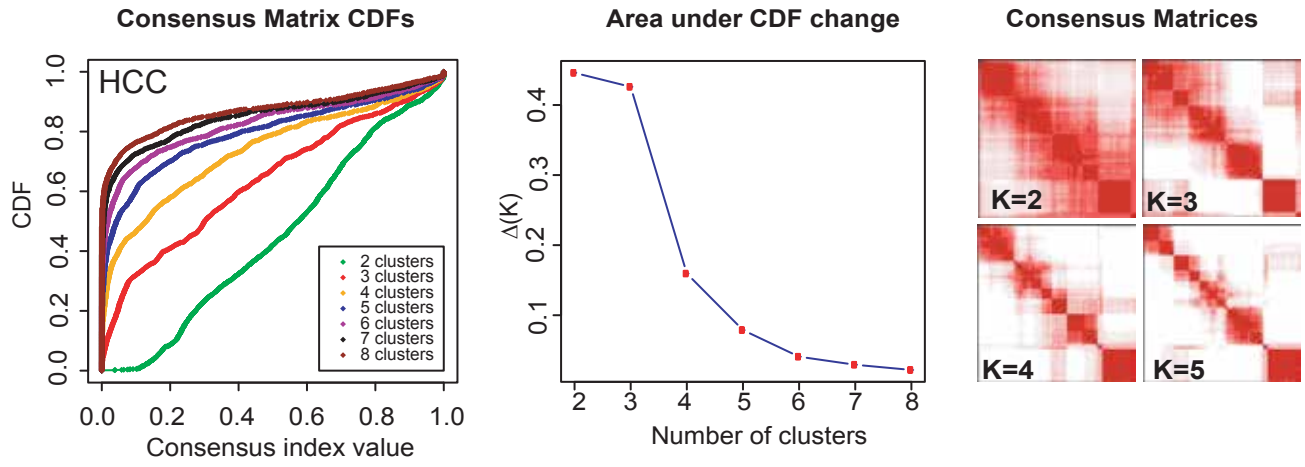
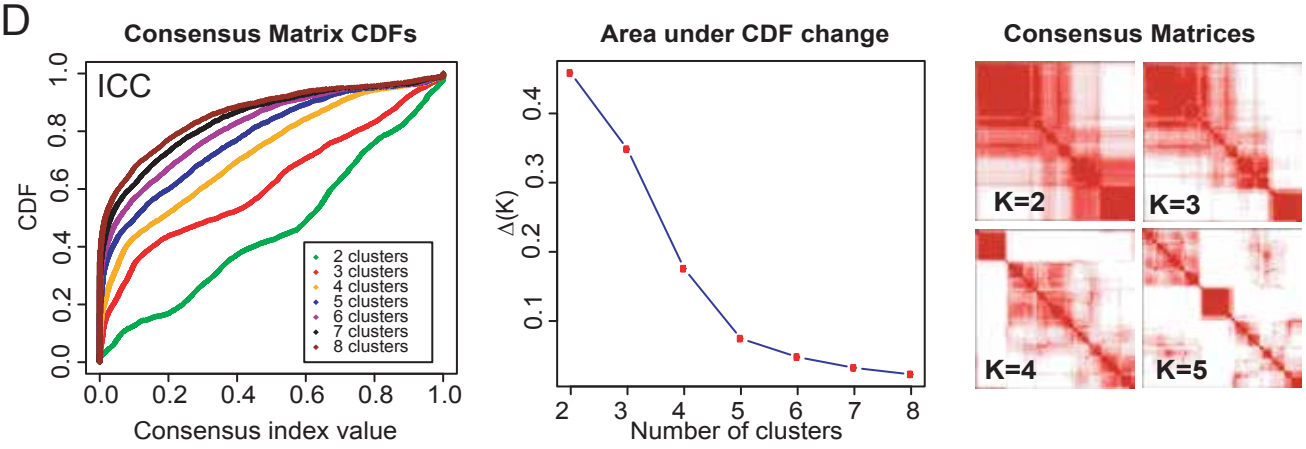

A
# of Thai patients
0
50
100
150
HCC
(69)
ICC
(130)
# of patients
0
100
200
300
400
Chinese HCC
(247)
TCGA HCC
(314)
European ICC
(104)
Japanese ICC
(182)
TIGER LC Cohorts
Validation cohorts
Transcriptome
Consensus clustering
Subclass mapping
Survival analysis
Pathway analysis
Common molecular subtypes
Subtype-specific signatures
Exome Seq
Oncovar v4
Alignment
Variant analysis
Subtype-related cancer driver
Copy Number Variation
Segmentation
Correlation; IPA
Subtype-related cancer driver
Metabolome
Correlation; IPA
Hierachical Clustering
Subtype-specific metabolic pathways


B
HCC Total (n=69)
Transcriptome (62)
Exome Seq (68)
CNV (62)
Metabolome (63)
TMA (69)
ICC Total (n=130)
Transcriptome (91)
Exome Seq (129)
CNV (90)
Metabolome (125)
TMA (130)

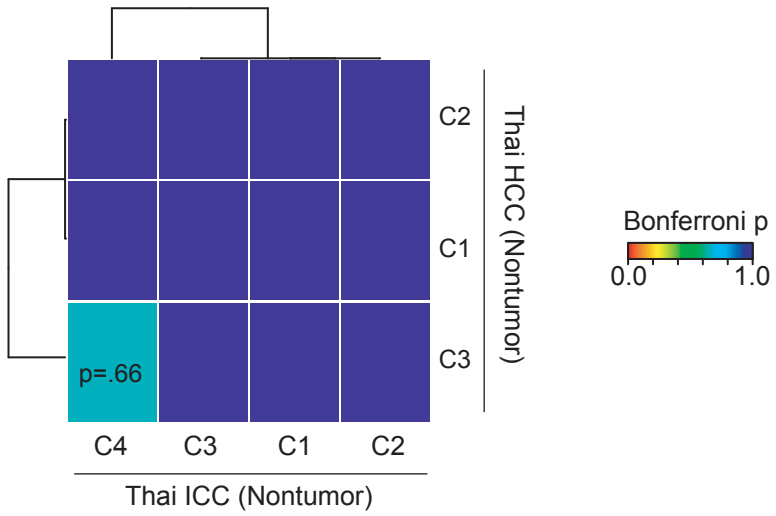
C


Consensus Matrix CDFs
HCC
CDF
Consensus index value
2 clusters
3 clusters
4 clusters
5 clusters
6 clusters
7 clusters
8 clusters
Area under CDF change
Δ(K)
Number of clusters
Consensus Matrices
K=2
K=3
K=4
K=5

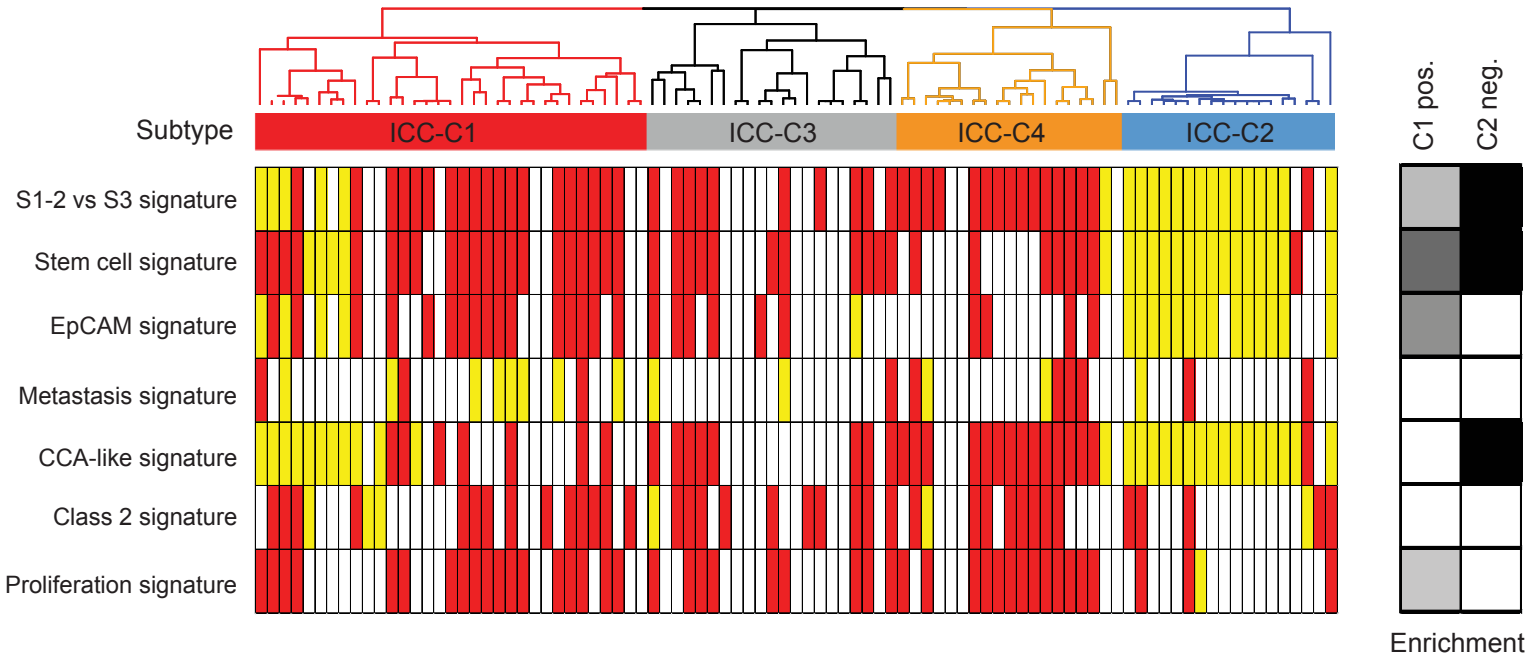
D


Consensus Matrix CDFs
ICC
CDF
Consensus index value
2 clusters
3 clusters
4 clusters
5 clusters
6 clusters
7 clusters
8 clusters
Area under CDF change
Δ(K)
Number of clusters
Consensus Matrices
K=2
K=3
K=4
K=5

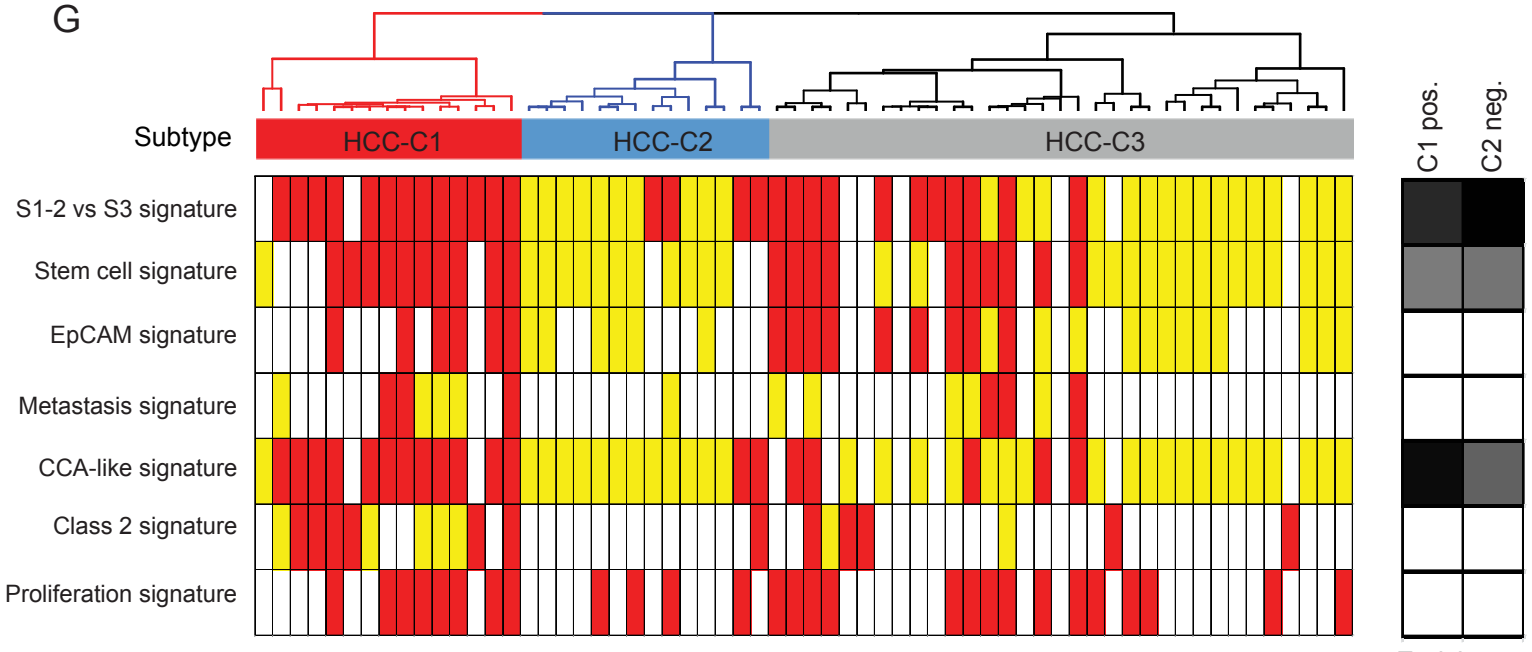
E


C2
C1
C3
Thai HCC (Nontumor)
p=.66
C4
C3
C1
C2
Thai ICC (Nontumor)
Bonferroni p
0.0
1.0

F



G



Positive for signature

Negative for signature

No correlation (FDR>0.05)

Enrichment
-log10 (adjusted p value)
7
0

Figure S1, related to Figure 1 and Figure 2. Study schema, Consensus clustering and Subclass Mapping of the Thai cohort and comparison of Thai ICC and HCC tumor subtypes to published signatures. (A) A schematic shows the cohorts used in this study and the related omics and bioinformatics pipelines. The number of samples are indicated in parentheses (B) The number of samples in each cohort and the number of samples used for each pipeline are indicated in parentheses. Results of consensus clustering of HCC (C) and ICC (D). Left panel: Empirical cumulative distribution function (CDF) plots corresponding to the consensus matrices in the range of cluster number (K) = 2-8, middle panel: Corresponding change in area under CDF, right panel: Consensus matrices corresponding to a number K of clusters ranging between 2 and 5. Results for ICC or HCC samples are shown on the top or bottom row, respectively. (E) Subclass mapping of nontumor tissue specimens of Thai HCC and ICC is shown. Significant relationships between clusters are represented by Bonferroni p values from 0 to 1. (F) A comparison between Thai ICC subtypes (n=91) and the published signatures indicated on the y-axis is shown. Each column represents a tumor sample. Positivity or negativity of each signature is represented by red bars or yellow bars, respectively. Cases with FDR >0.05 are indicated in white. The enrichment p value based on a one-sided Fisher's exact test is shown to determine the presence (positive: pos) or absence (negative: neg) of each signature among the C1 or C2 subtypes. (G) A comparison between Thai HCC subtypes (n=62) and the published signatures indicated on the y-axis is shown.

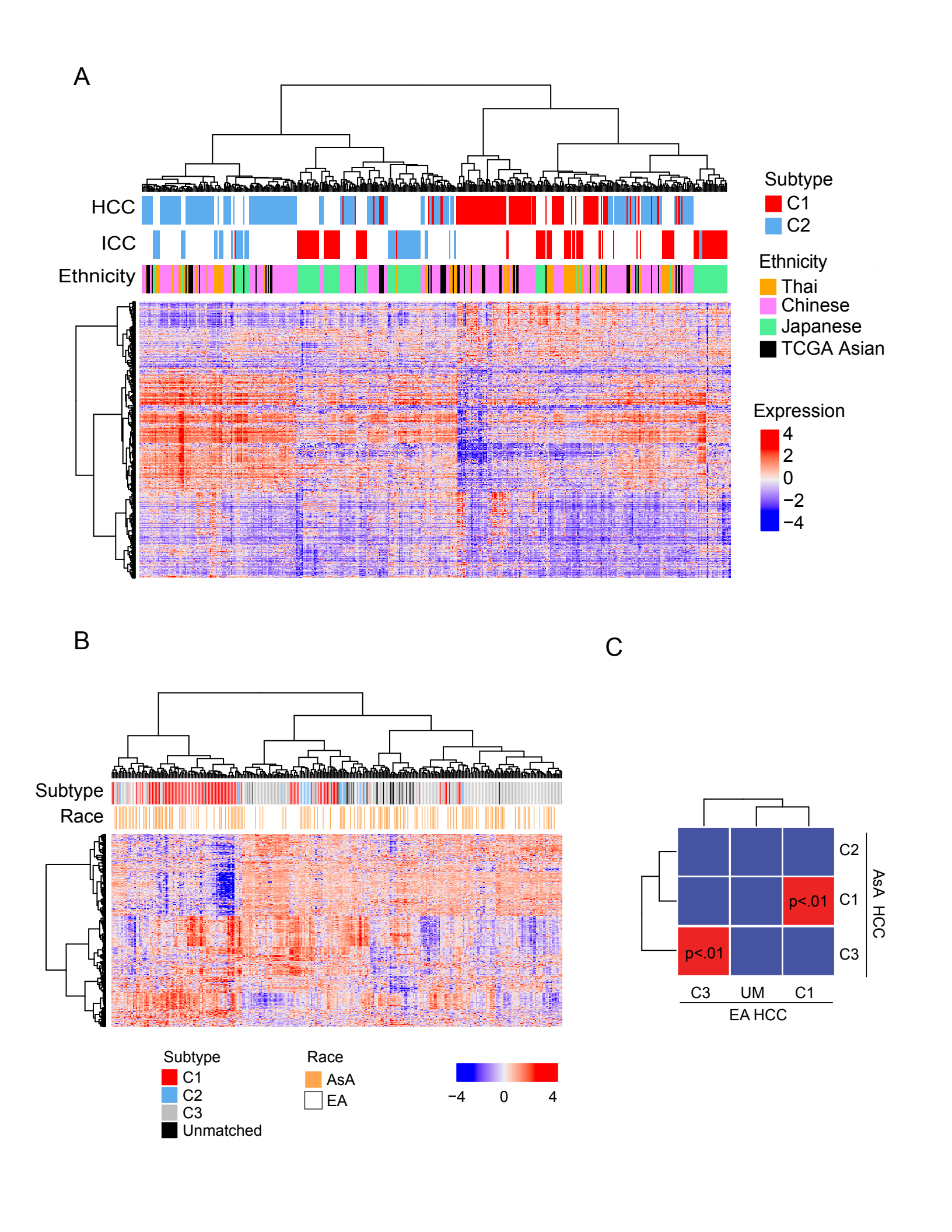
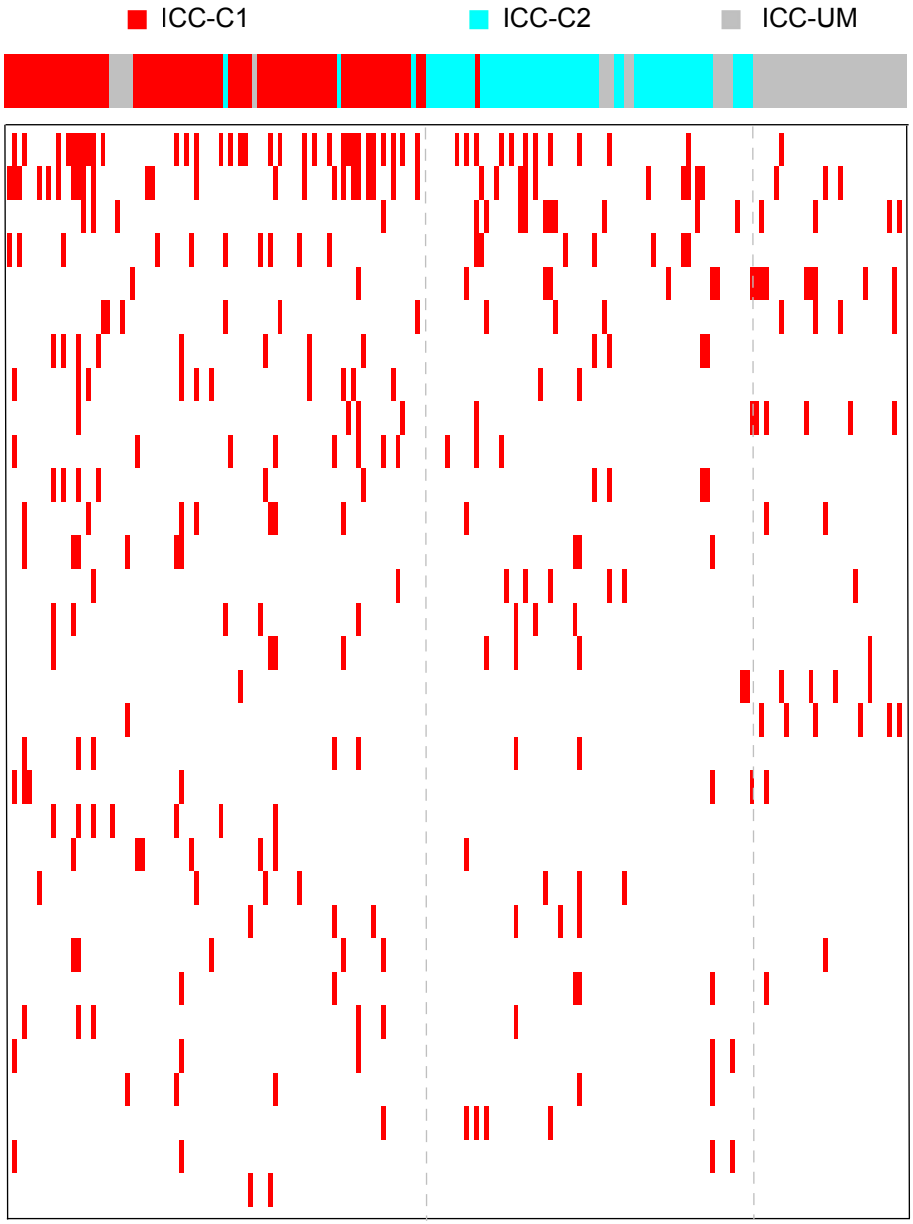

A
HCC
ICC
Ethnicity
Subtype
C1
C2
Ethnicity
Thai
Chinese
Japanese
TCGA Asian
Expression
4
2
0
−2
−4
B
Subtype
Race
Subtype
C1
C2
C3
Unmatched
Race
AsA
EA
−4
0
4
C
C2
C1
C3
p<.01
p<.01
AsA HCC
C3
UM
C1
EA HCC

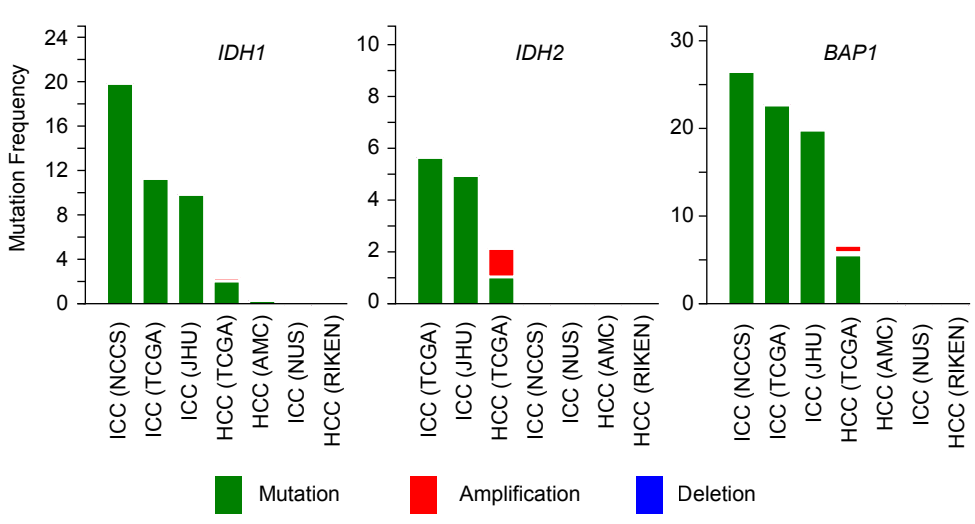
Figure S2, related to Figure 3. The C1 and C2 subtypes of HCC and ICC cluster together among cohorts of Asian descent and comparisons of subtypes to HCC patients of Caucasian decent in TCGA. (A) A heatmap of HCC and ICC C1 and C2 subtypes is shown based on hierarchical clustering of the most variable genes between C1 and C2 (Pearson distance, Ward linkage). The x-axis represents HCC or ICC C1 and C2 subtypes (red bars: cluster C1; blue: cluster C2). Samples are represented in columns, grouped by the dendrogram and genes are represented in rows. Z-scored gene expression is shown from -4 to 4. Samples from Thai (yellow bars), Chinese (pink bars), Japanese (green bars) or Asian TCGA (black bars) cohorts are shown. (B) A heatmap of HCC subtypes is shown based on hierarchical clustering of the most differentially ranked genes between C1 and C2 subtypes (n=637) among HCC patients in TCGA. The x-axis represents HCC subtypes (red bars: cluster C1; blue: cluster C2; grey: C3; black: unmatched) and the patient race (yellow: Asian; white: Caucasian). Samples are represented in columns, grouped by the dendrogram and genes are represented in rows. Gene expression are shown in log2 from -4 to 4. (C) Subclass Mapping of Asian or Caucasian subtypes of HCC patients in TCGA is shown. Significant relationships between clusters are represented by Bonferroni adjusted p values. Significant associations between clusters are shown in red with $p<0.05$.

A

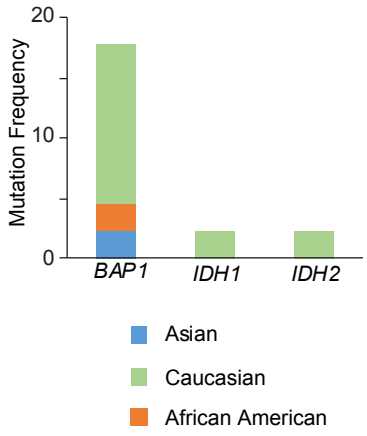


	%		
	C1	C2	UM
TP53	43	15	4
KRAS	30	17	7
ARID1A	5	14	13
SMAD4	13	12	0
BAP1	1	7	26
PIK3CA	5	5	13
MDM2	10	5	2
MYC	13	3	0
FGFR2	6	0	13
GNAS	12	3	0
YEATS4	8	5	2
HSF1	9	2	4
EGFR	6	3	4
ARID2	3	5	7
FGFR1	5	7	0
FGFR3	5	5	2
NRAS	1	3	9
IDH1	0	0	15
PRKCI	6	3	0
DDR2	5	0	7
WNT5B	8	0	2
CDKN2A	8	2	0
ELF3	5	3	2
CCND3	4	5	0
IKBKB	6	0	2
AKT3	3	3	4
ECT2	6	2	0
FGF19	4	2	2
GNA12	3	2	4
RPL22	3	5	0
CCND1	3	2	2
CCNE1	3	0	0

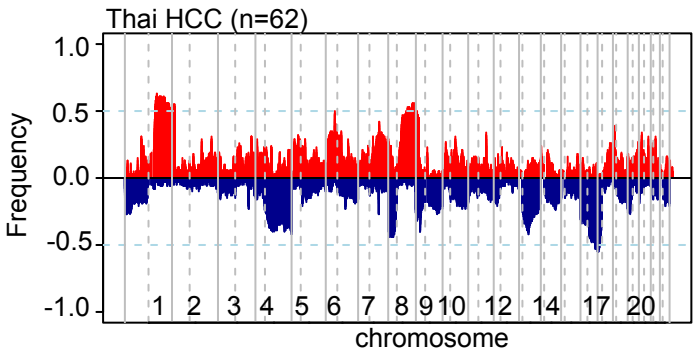
B

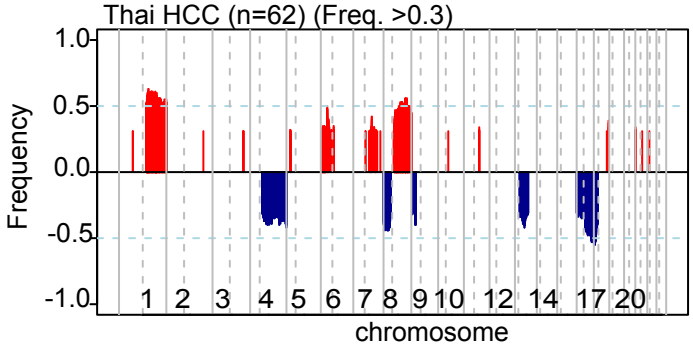


C

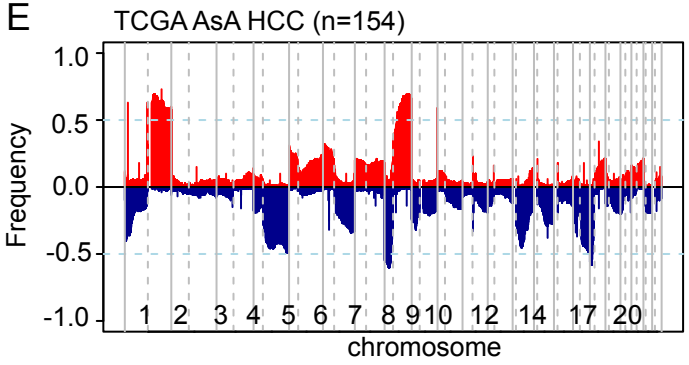


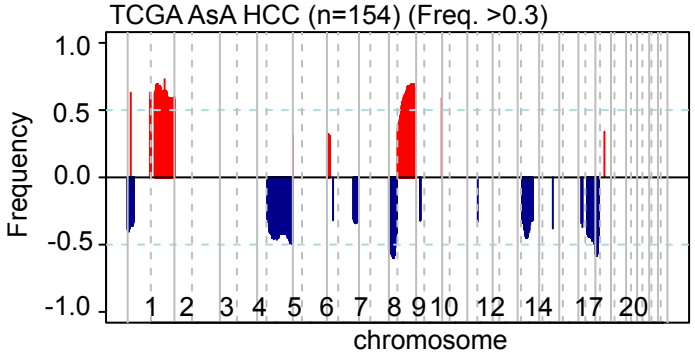
D



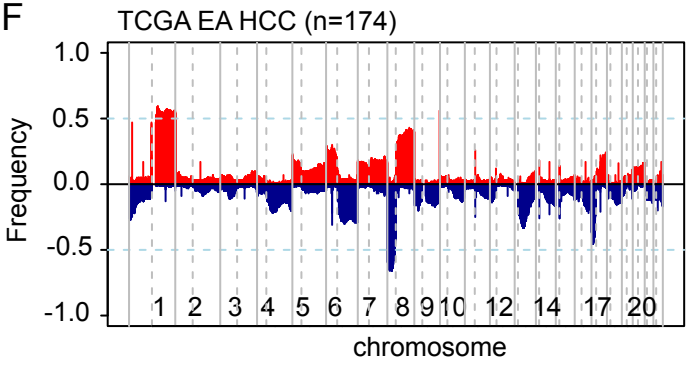


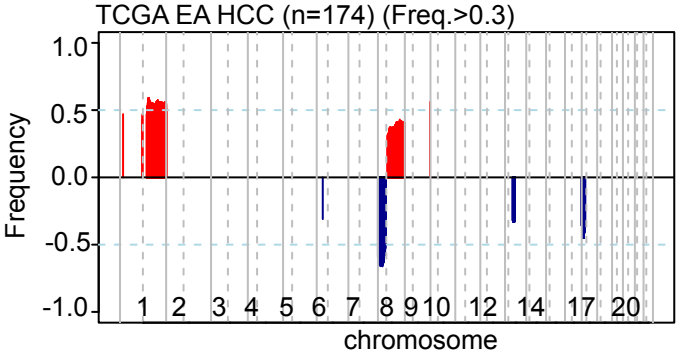
E





F



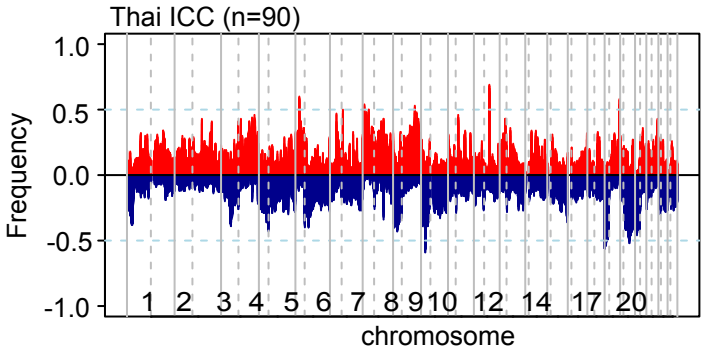


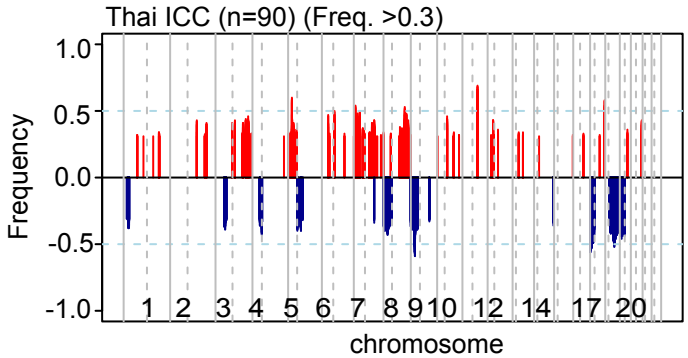
Shared
SCNA: 1q, 6p, 7q, 8q (Gain); 4q, 8p, 9p, 13q, 16, 17p (loss)

1q, 6p, 8q (Gain); 1p, 4q, 6q, 8p, 13q, 16, 17p (loss)

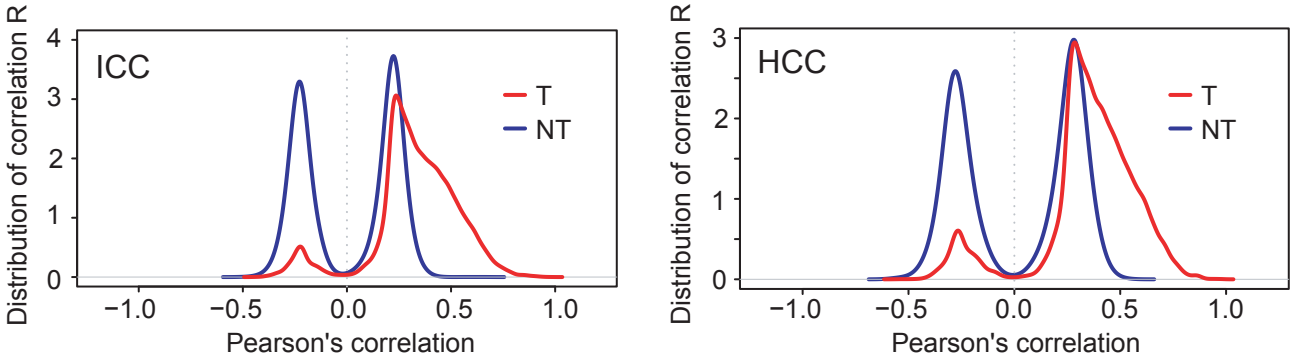
1q, 8q (Gain); 8p, 13q, 17p (loss)

G

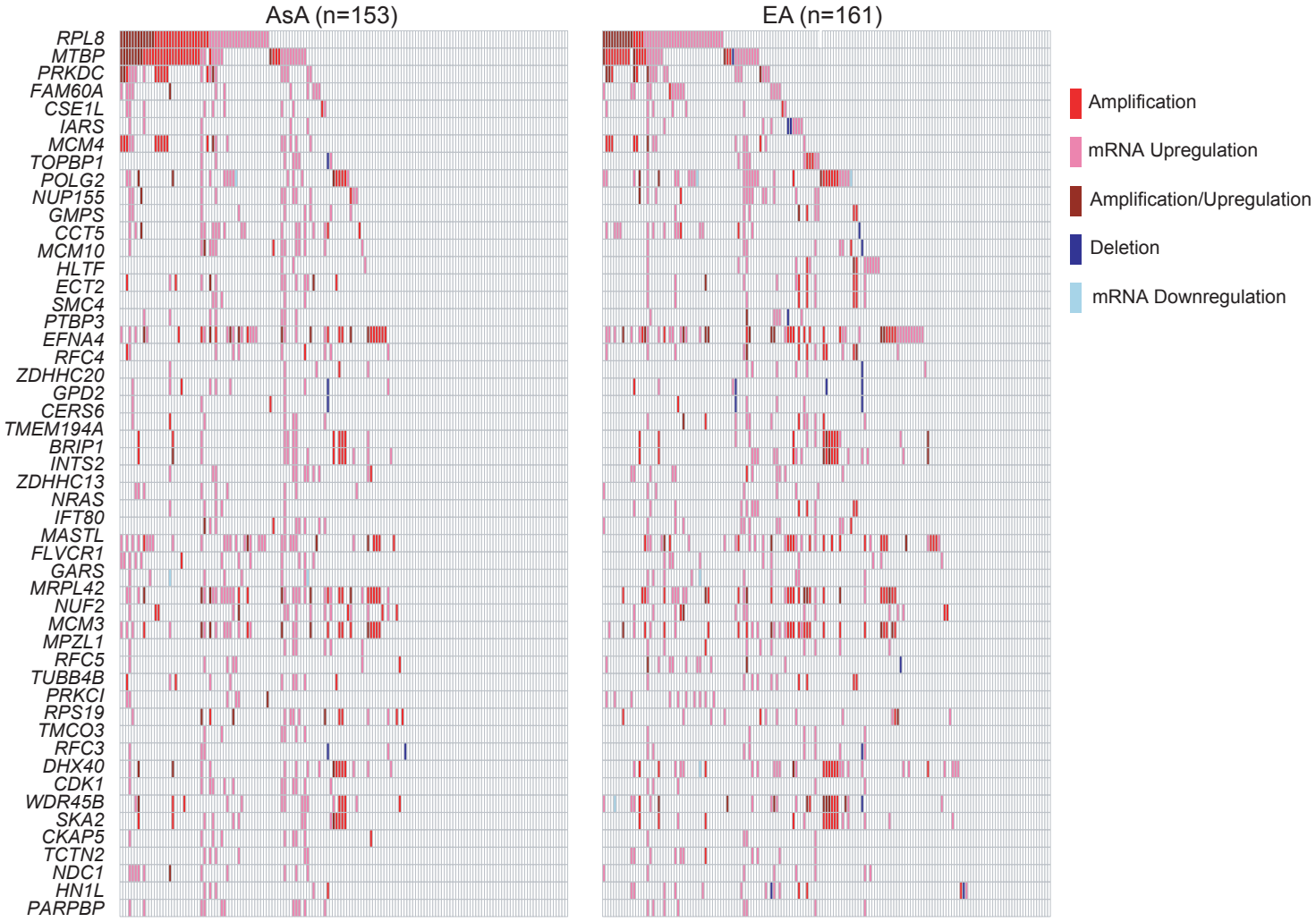




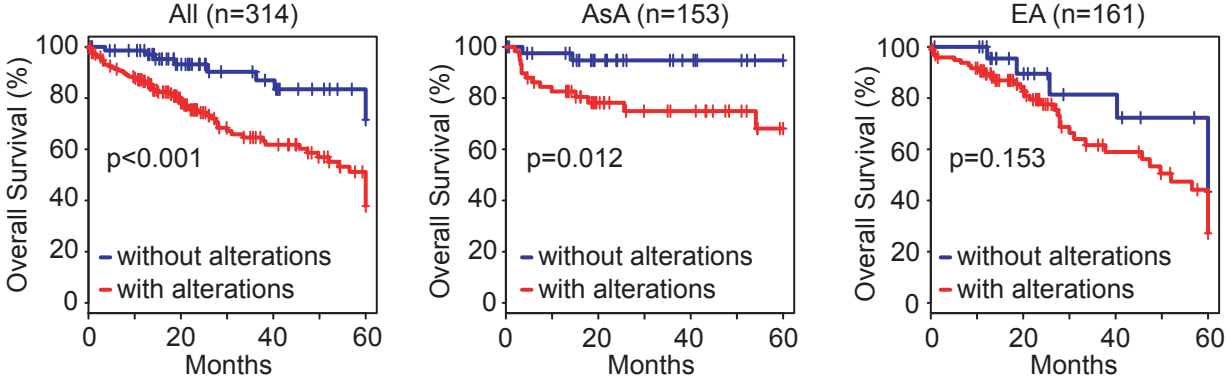
H

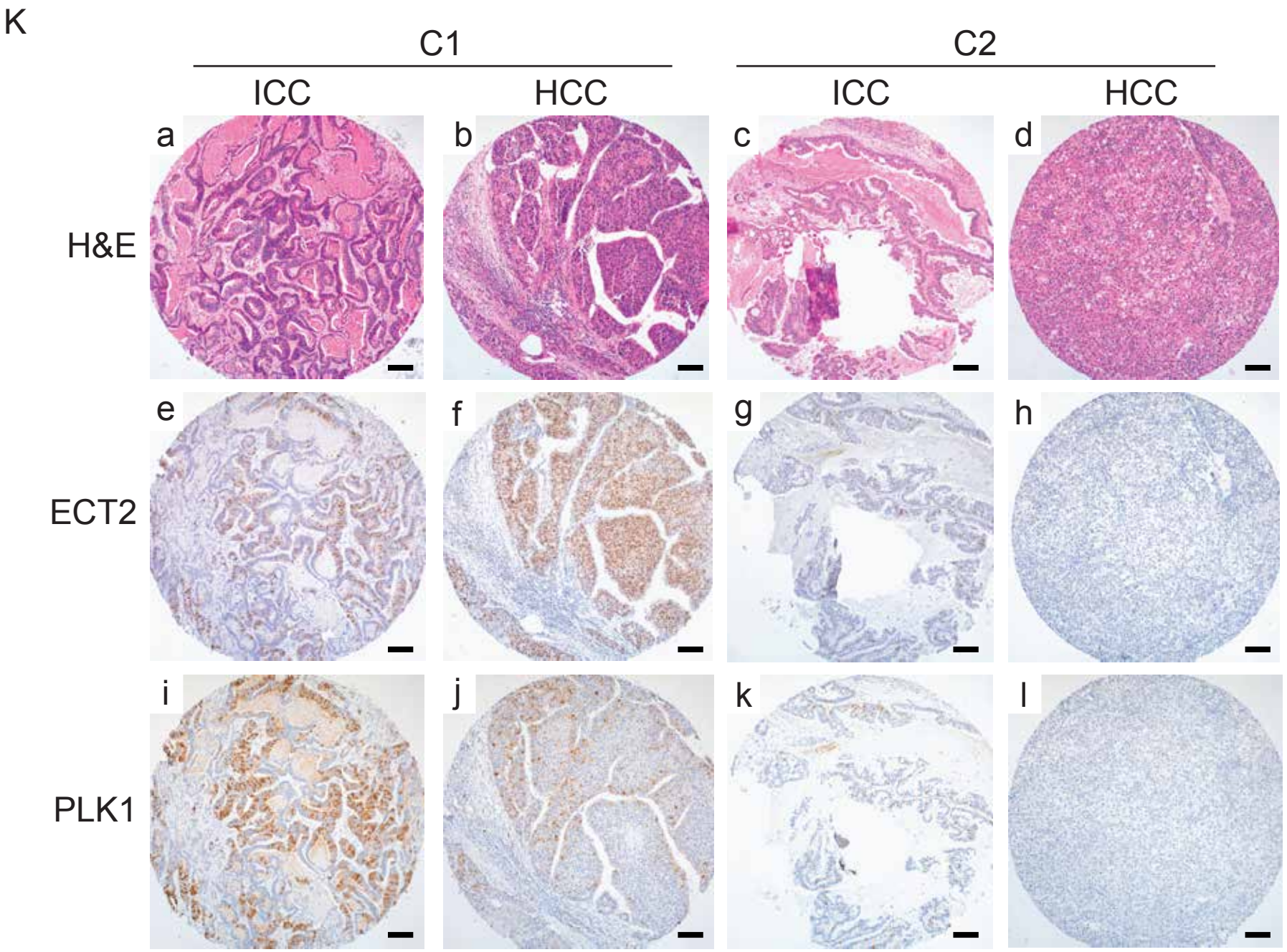

ICC
HCC
T
NT
Distribution of correlation R
Pearson's correlation

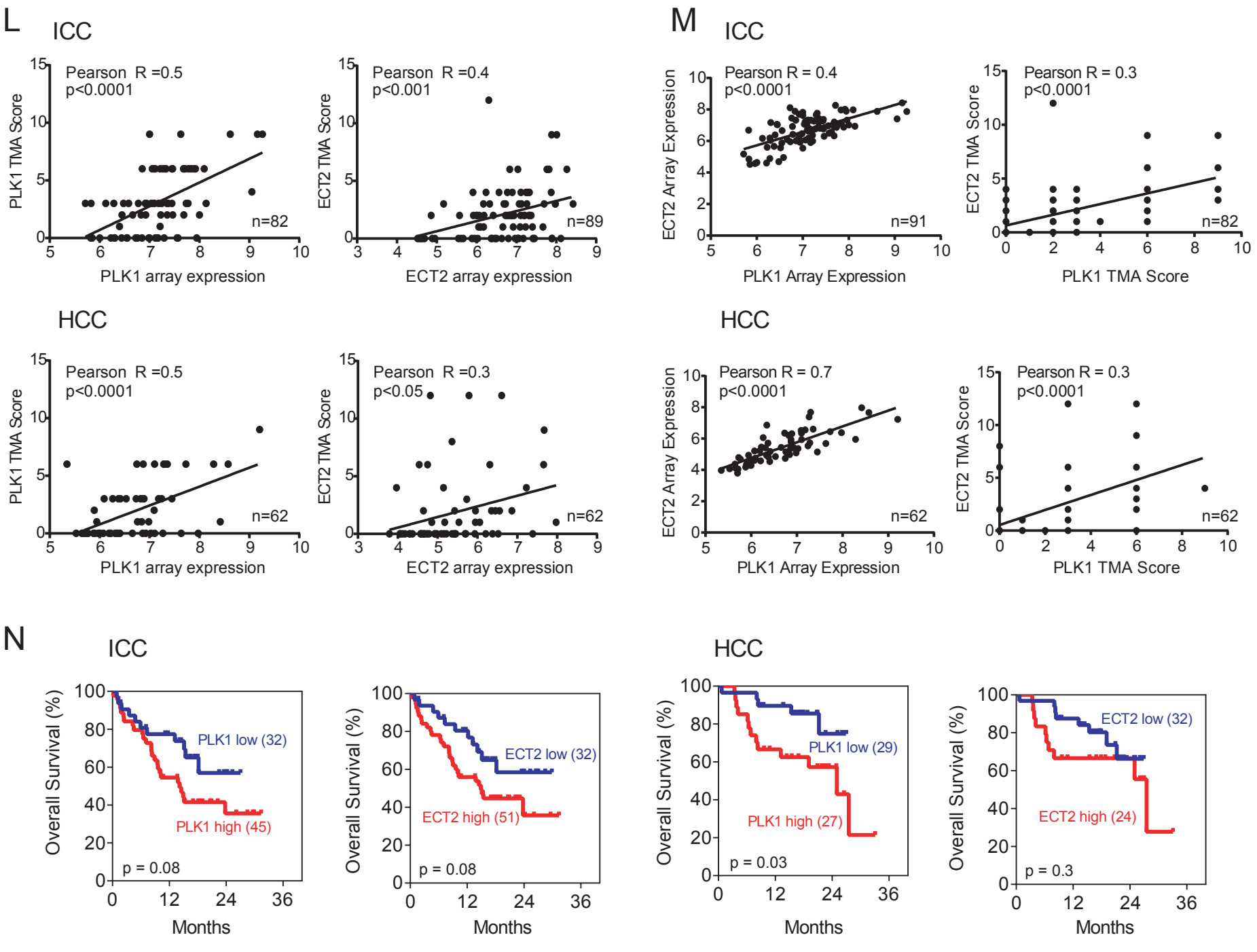
I


AsA (n=153)
EA (n=161)
RPL8
MTBP
PRKDC
FAM60A
CSE1L
IARS
MCM4
TOPBP1
POLG2
NUP155
GMPS
CCT5
MCM10
HLTF
ECT2
SMC4
PTBP3
EFNA4
RFC4
ZDHHC20
GPD2
CERS6
TMEM194A
BRIP1
INTS2
ZDHHC13
NRAS
IFT80
MASTL
FLVCR1
GARS
MRPL42
NUF2
MCM3
MPZL1
RFC5
TUBB4B
PRKCI
RPS19
TMCO3
RFC3
DHX40
CDK1
WDR45B
SKA2
CKAP5
TCTN2
NDC1
HN1L
PARPBP
Amplification
mRNA Upregulation
Amplification/Upregulation
Deletion
mRNA Downregulation

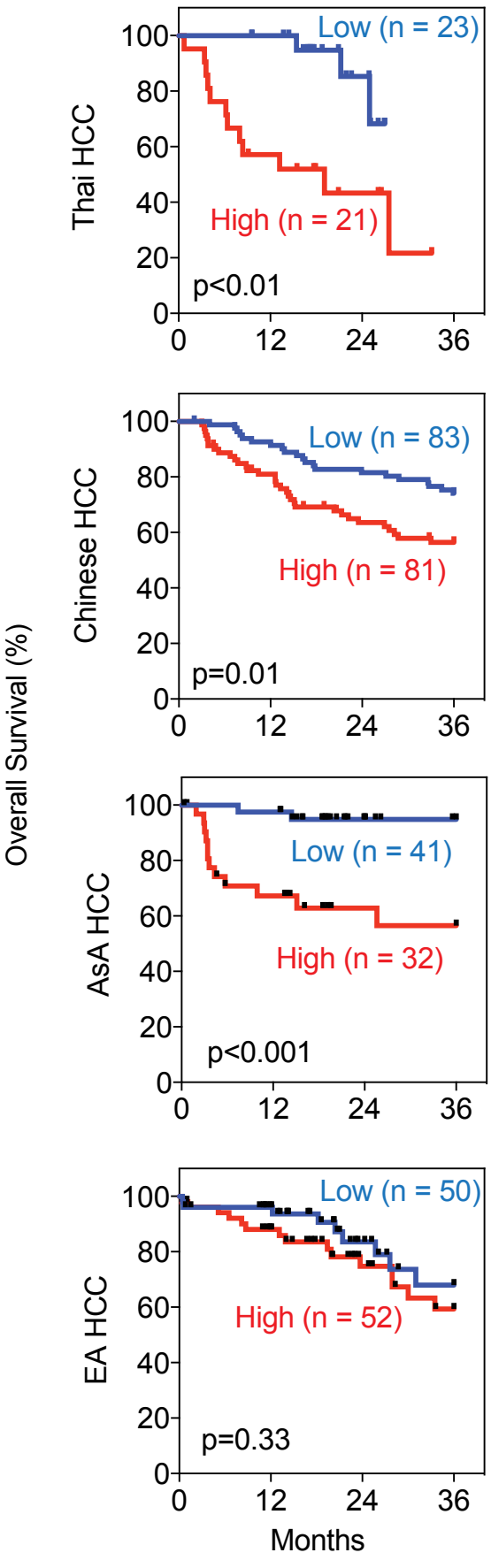
J


All (n=314)
AsA (n=153)
EA (n=161)
Overall Survival (%)
p<0.001
p=0.012
p=0.153
without alterations
with alterations
Months


K
C1
C2
ICC
HCC
ICC
HCC
H&E
ECT2
PLK1
a
b
c
d
e
f
g
h
i
j
k
l


L
ICC
Pearson R =0.5
p<0.0001
PLK1 TMA Score
PLK1 array expression
n=82
Pearson R =0.4
p<0.001
ECT2 TMA Score
ECT2 array expression
n=89
HCC
Pearson R =0.5
p<0.0001
PLK1 TMA Score
PLK1 array expression
n=62
Pearson R =0.3
p<0.05
ECT2 TMA Score
ECT2 array expression
n=62
M
ICC
Pearson R = 0.4
p<0.0001
ECT2 Array Expression
PLK1 Array Expression
n=91
Pearson R = 0.3
p<0.0001
ECT2 TMA Score
PLK1 TMA Score
n=82
HCC
Pearson R = 0.7
p<0.0001
ECT2 Array Expression
PLK1 Array Expression
n=62
Pearson R = 0.3
p<0.0001
ECT2 TMA Score
PLK1 TMA Score
n=62
N
ICC
Overall Survival (%)
PLK1 low (32)
PLK1 high (45)
p = 0.08
Months
ECT2 low (32)
ECT2 high (51)
p = 0.08
Months
HCC
Overall Survival (%)
PLK1 low (29)
PLK1 high (27)
p = 0.03
Months
ECT2 low (32)
ECT2 high (24)
p = 0.3
Months

O


Overall Survival (%)
Thai HCC
Low (n = 23)
High (n = 21)
p<0.01
Chinese HCC
Low (n = 83)
High (n = 81)
p=0.01
AsA HCC
Low (n = 41)
High (n = 32)
p<0.001
EA HCC
Low (n = 50)
High (n = 52)
p=0.33
0
20
40
60
80
100
0
12
24
36
Months

P

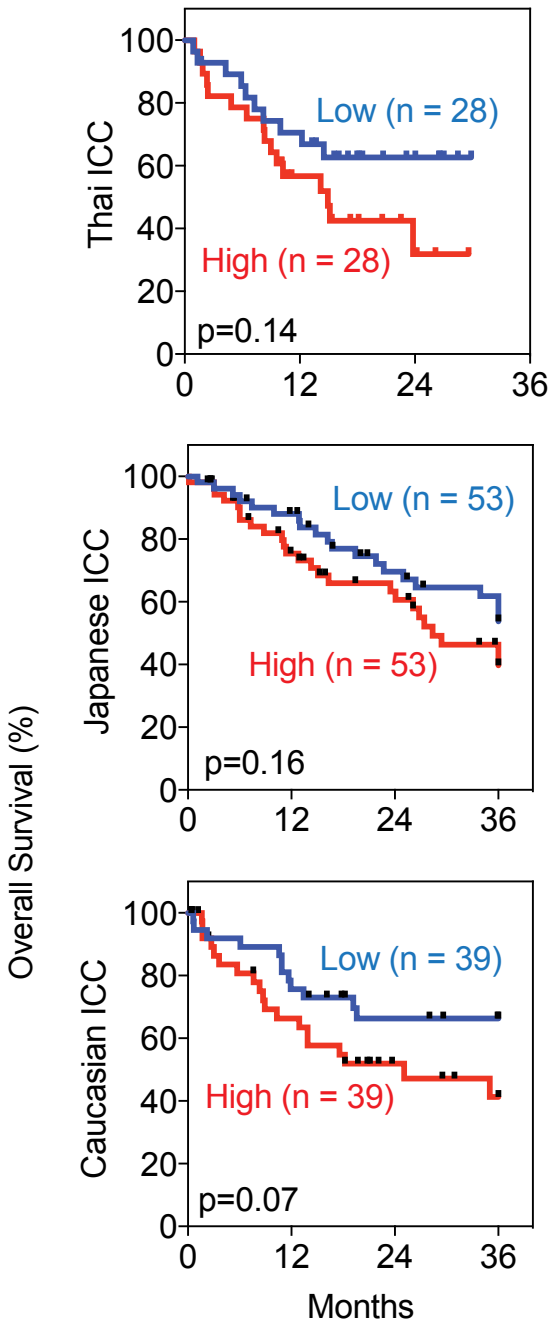
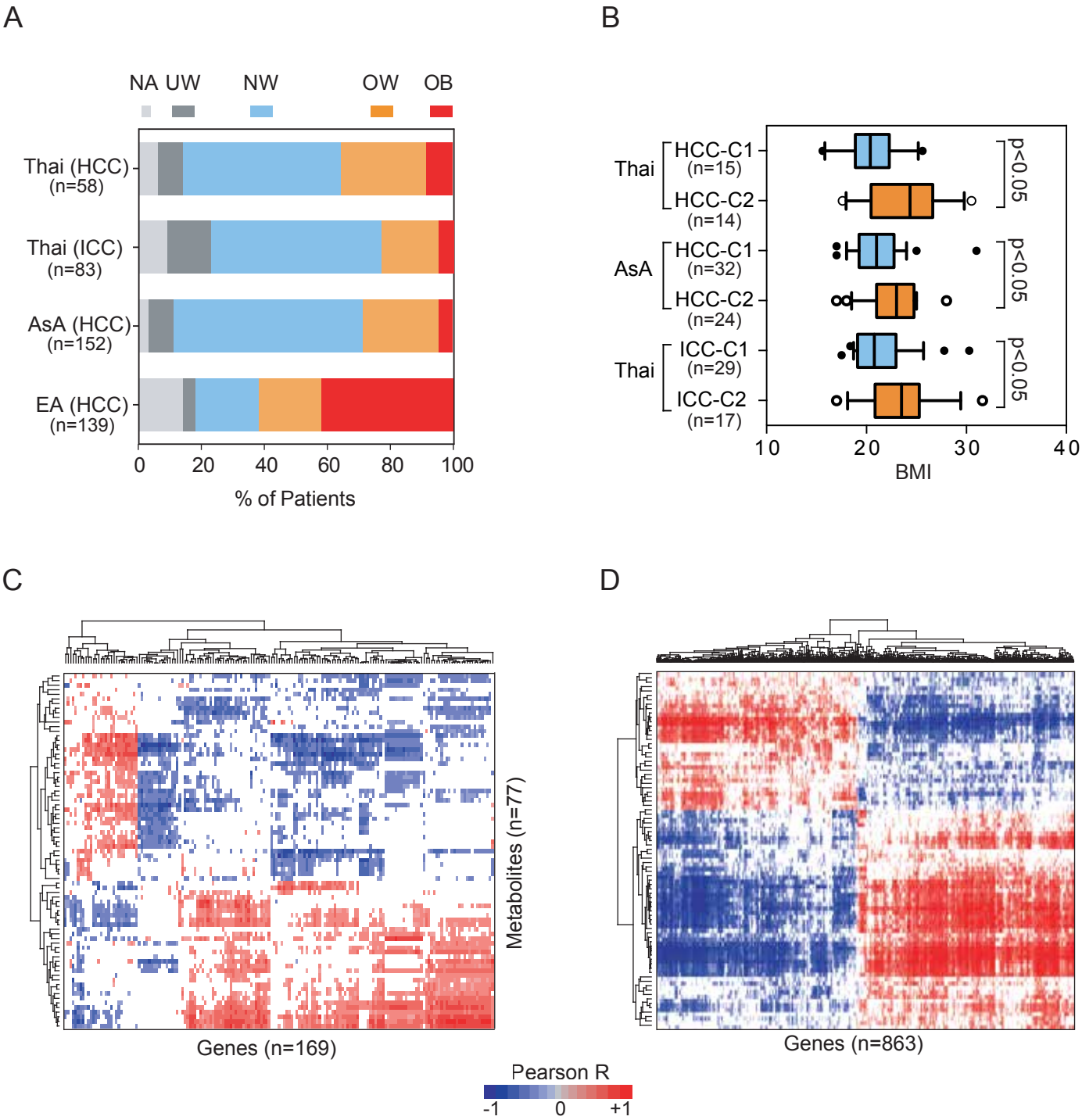

Overall Survival (%)
Thai ICC
Low (n = 28)
High (n = 28)
p=0.14
Japanese ICC
Low (n = 53)
High (n = 53)
p=0.16
Caucasian ICC
Low (n = 39)
High (n = 39)
p=0.07
0
20
40
60
80
100
0
12
24
36
Months

Figure S3, related to Figure 6. Relationship between ICC subtypes with commonly mutated genes in the Japanese cohort, somatic copy number alterations (SCNA), correlation with gene expression in ICC or HCC and PLK1 and ECT2 expression and their association with prognosis. (A) The frequency of commonly mutated genes in the C1 (n=77) or C2 subtype (n=59) of 182 Japanese ICC tumors as well as an unmatched group (UM, n=46) are shown. The percent of cases of each subtype with mutation are shown in the right panel and the number of cases in each subtype is indicated in parentheses. (B) The mutation frequency of IDH1, IDH2 or BAP1 is shown among the indicated ICC or HCC cohorts on the y-axis. The type of alteration (green: mutation; red: amplification; blue: deletion) is shown. (C The mutation frequency of IDH1, IDH2 or BAP1 is shown among ICC patients in TCGA stratified by race (Asian, Caucasian or Black). (D) The frequency of chromosomal aberrations is shown for HCC in the Thai cohort (upper panel) or chromosomal aberrations with frequency >30% (lower panel). (E) The same results are shown for Asian patient in TCGA, (F) Caucasian patients in TCGA and (G) ICC patients in the Thai cohort. Shared regions of copy number gain or loss are indicated below the frequency plots and the number of samples are indicated in parentheses. (H) The distribution of Pearson's correlation between SCNA and gene expression is shown in either tumor (T) or nontumor (NT) specimens of ICC (left panel) or HCC (right panel). (I) The relationship between chromosomal alteration and gene expression of 51 driver genes in the TCGA HCC cohort (AsA, n=153; EA, n=161). Bars indicate amplification (red), deletion (blue); increased expression (pink), decreased expression (light blue), both increased gene amplification and expression (dark red). (J) Kaplan-Meier plots of all 314 HCC cases, 153 AsA cases or 161 EA cases with or without alterations are shown. (K) Representative images of an ICC-C1 (panels a, e, i) or an HCC-C1 (panels b, f, j) and an ICC-C2 (panels c, g, k), or a HCC-C2 case (panel d, h, l) are shown at 200x magnification based on immunohistochemical staining for ECT2 (panels e-h) or PLK1 (panels i-l). H&E images of the same regions of each case are shown (panels a-d). Scale bars represent 10 μm. (L) Correlation of PLK1 (left panels) or ECT2 (right panels) array expression and TMA score in ICC cases (upper panel) or HCC cases (lower panel) is shown. (M) Correlation of PLK1 and ECT2 array expression or TMA score in ICC or HCC cases is shown. (N) Kaplan-Meier survival analysis of ICC cases (left panels) or HCC cases (right panels) is shown based on a median cutoff of PLK or ECT2 with the number of samples indicated in parentheses. (O) Kaplan-Meier survival analysis of HCC among Thai, Chinese, Asian TCGA or Caucasian TCGA cohorts based on the median mRNA expression of both PLK1 and ECT2 is shown with log-rank p-value. The number of samples is indicated in parentheses. (P) Similar Kaplan-Meier survival analysis of ICC among Thai, Japanese or Caucasian patients is shown.


A
NA UW NW OW OB
Thai (HCC) (n=58)
Thai (ICC) (n=83)
AsA (HCC) (n=152)
EA (HCC) (n=139)
0 20 40 60 80 100
% of Patients
B
Thai
HCC-C1 (n=15)
HCC-C2 (n=14)
AsA
HCC-C1 (n=32)
HCC-C2 (n=24)
Thai
ICC-C1 (n=29)
ICC-C2 (n=17)
p>0.05
p>0.05
p>0.05
10 20 30 40
BMI
C
Metabolites (n=77)
Genes (n=169)
D
Metabolites (n=81)
Genes (n=863)
Pearson R
-1 0 +1

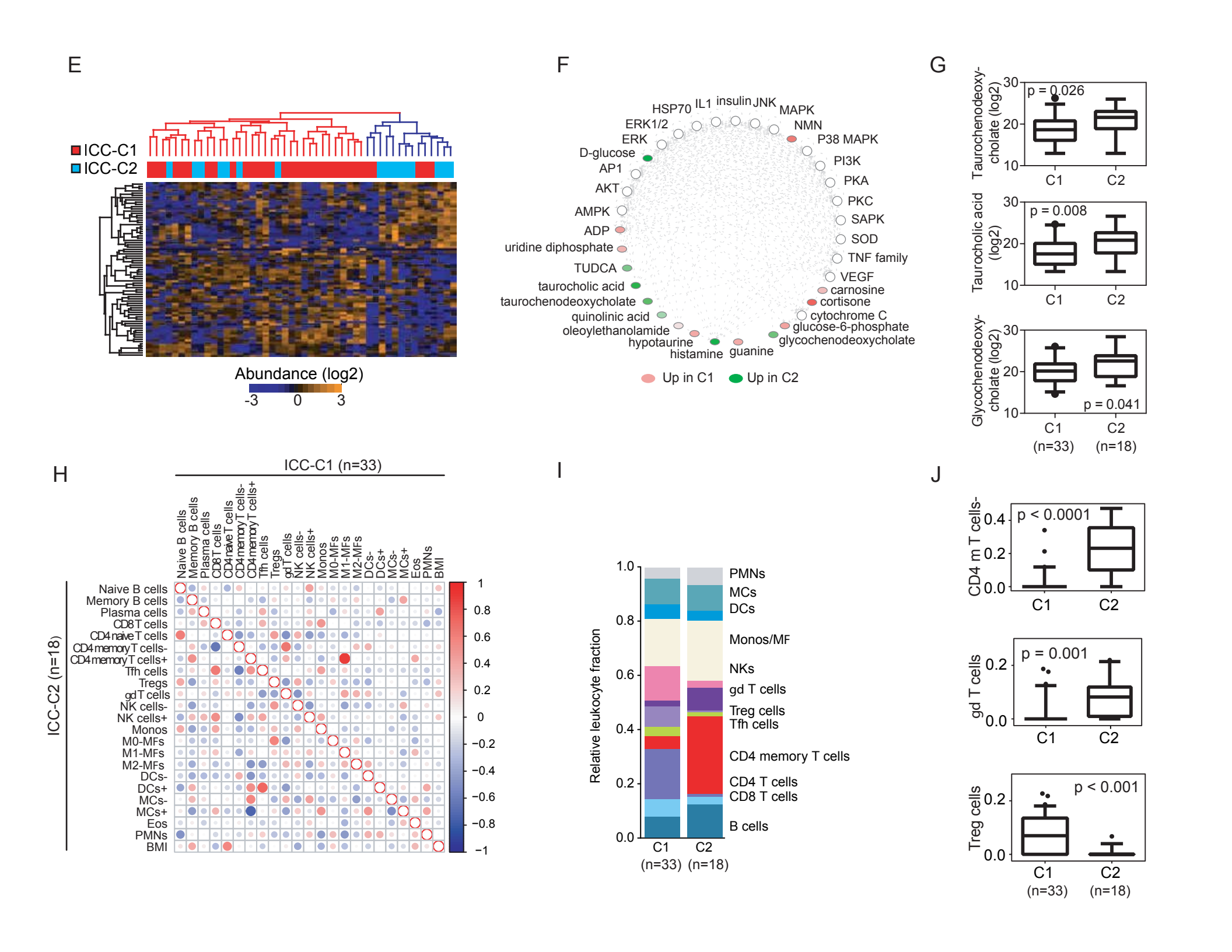

E
ICC-C1
ICC-C2
Abundance (log2)
-3
0
3
F
HSP70
IL1
insulin
JNK
MAPK
NMN
P38 MAPK
PI3K
PKA
PKC
SAPK
SOD
TNF family
VEGF
carnosine
cortisone
cytochrome C
glucose-6-phosphate
glycochenodeoxycholate
guanine
histamine
hypotaurine
oleoylethanolamide
quinolinic acid
taurochenodeoxycholate
taurocholic acid
TUDCA
uridine diphosphate
ADP
AMPK
AKT
AP1
D-glucose
ERK
ERK1/2
Up in C1
Up in C2
G
Taurochenodeoxy-cholate (log2)
p = 0.026
Taurocholic acid (log2)
p = 0.008
Glycochenodeoxy-cholate (log2)
p = 0.041
C1
C2
(n=33)
(n=18)
H
ICC-C1 (n=33)
ICC-C2 (n=18)
Naive B cells
Memory B cells
Plasma cells
CD8 T cells
CD4 naive T cells
CD4 memory T cells-
CD4 memory T cells+
Tfh cells
Tregs
gd T cells
NK cells-
NK cells+
Monos
M0-MFs
M1-MFs
M2-MFs
DCs-
DCs+
MCs-
MCs+
Eos
PMNs
BMI
I
Relative leukocyte fraction
PMNs
MCs
DCs
Monos/MF
NKs
gd T cells
Treg cells
Tfh cells
CD4 memory T cells
CD4 T cells
CD8 T cells
B cells
J
CD4 m T cells-
p < 0.0001
gd T cells
p = 0.001
Treg cells
p < 0.001

Figure S4, related to Figure 7. BMI status and correlation of metabolite and gene expression are correlated in ICC and HCC identify altered bile-acid metabolism and inflammation in the ICC C1 and C2 subtype. (A) The percent of patients in body mass index (BMI) categories among the Thai HCC, Thai ICC, Asian American (AsA) HCC and European American (EA) HCC cases is shown. BMI is categorized as underweight (UW), normal weight (NW), overweight (OW), obese (OB) or not available (NA) according to WHO criteria adjusted for the Asian population (underweight, BMI <18.5, normal weight, BMI = 18.5–23.9; overweight, BMI =24-27.9, obese, BMI >28) or BMI categories for European Americans: underweight, BMI <18.5, normal weight, BMI = 18.5–24.9; overweight, BMI =25-29.9, obese, BMI >30. The number of cases in each subtype is indicated in parentheses. (B) Box plots of BMI values with first quartile, median and third quartile (bottom box, middle line and top box, respectively) are shown for C1 and C2 subtypes of Thai HCC, Asian American HCC and Thai ICC cases. Whiskers represent 10-90th percentile with dots as outliers. The number of cases in each subtype is indicated in parentheses. (C) A heatmap represents the correlation of metabolite abundance and gene expression in ICC samples. Red and blue bars indicate positive or negative correlation, respectively, based on the Pearson R value from -1 to 1. (D) A heatmap is shown representing the global correlation of metabolite abundance and gene expression in HCC samples. The number of cases in each subtype is indicated in parentheses. (E) Hierarchical clustering of 77 metabolites separates ICC-C1 (red bar) and C2 (blue bar) HCC cases (n=51). Samples are represented in columns, and metabolites are represented in rows. Metabolite abundance is represented in log2. (F) Ingenuity Pathway Analysis of the highly concordant metabolite/gene network is shown. Upregulated metabolites in C1 subtype or the C2 subtype are noted in pink or green, respectively. (G) Box-plots of the abundance of three representative bile-acid-related metabolites in C1 (n=33) and C2 (n=18) ICC samples are shown as first quartile, median and third quartile (bottom box, middle line and top box, respectively) with Student's t-test p values. Whiskers represent 10-90th percentile with dots as outliers. (H) CIBERSORT analysis of the ICC C1 versus the ICC C2 subtype is shown. High or low associations between cell types are shown on a scale from red to blue (1 to -1). The size of circles indicates the significance of the association, with larger circles representing higher significance. (I) The relative fraction of leukocyte types associated with C1 and C2 are shown. (J) Box-plots of the abundance of three leukocyte types in C1 and C2 ICC samples are shown as first quartile, median and third quartile (bottom box, middle line and top box, respectively) with Student's t-test p-values. Whiskers represent 10-90th percentile with dots as outliers.